Subject: Merging BR and WI in the Egypt 2000 survey
Posted by yanesko on Thu, 07 Jan 2016 16:25:50 GMT
View Forum Message <> Reply to Message

Dear DHS,

When I try to merge the BR and WI files in the Egypt_2000 survey, I get 0 matches. I was wondering if the WI file is actually not from this survey (since the number of the Phase in the file name is different)?

Here is the STATA code I use:

use EGBR42FL.DTA
gen whhid = substr(caseid, 1, 12)
merge m:1 whhid using EGWI41FL.DTA, keepusing(wlthind5) keep(match master)

I used the same code for other surveys and it worked fine.

Thank you in advance!

Yana

Subject: Re: Merging BR and WI in the Egypt 2000 survey
Posted by Bridgette-DHS on Tue, 12 Jan 2016 18:03:52 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

The fifth character in the file name is the phase (we are now in phase 7 of DHS) and the sixth character is the version, which is initially 0 and is incremented by 1 when there is some kind of updating of a file. The files EGBR42FL.dta and EGWI41FL.dta are in the same phase (4) and refer to the same survey, but have different version codes (2 and 1, respectively). It is not a problem that the versions are different. (Note: a subsequent survey within phase 4 would be identified with phase=4 and version=A, and then the update would have version=B, etc.)

The reason you cannot do the merge is that the household id code in the WI file, whhid, is str12, i.e. is a 12-character string. In the BR file, the households are identified by v001 (cluster) and v002 (household within cluster), both of which are numeric. It's a real nuisance when this problem comes up. You have to separate whhid into two separate strings and then convert those strings to numeric.

To see where the break is between v001 and v002 within whhid, just list whhid for the first few cases in the WI file and compare with a list of v001 and v002 in the BR file. I did this and I can see that whhid needs to be broken into 9 characters + 3 characters.

When you open the WI file, prior to doing the merge use these three lines:

```
gen v001=substr(whhid,1,9)
gen v002=substr(whhid,10,3)
destring v001 v002, replace
```

Then sort both files on v001 v002 and do the merge.